Technical Disclosure Commons

Defensive Publications Series

November 2019

Capture Camera Movement Compensation for Display of 3D Video in Virtual Reality

Anonymous

Capture Camera Movement Compensation for Display of 3D Video in Virtual Reality

ABSTRACT

Videos captured with smartphones or other handheld cameras can include shaking due to motion of the user's hand during capture. When such videos are transformed into 3D for rendering in a virtual reality environment, the shaking in the 2D video can cause the viewer of the VR scene to experience a roller coaster feeling where the virtual world shakes. This disclosure describes techniques to track points across multiple frames of a 2D video. When the video is transformed into 3D, a 3D coordinate is established for each pixel of the video. The tracked points are used as input to optimization techniques that determine rotation and/or translation operations to transform the coordinate system such that the 3D world is stabilized.

KEYWORDS

- Virtual reality
- Roller coaster effect
- 3D video
- Video stabilization
- Camera shaking
- Camera movement

BACKGROUND

Users often capture videos using a device such as a smartphone or other handheld camera. The inadvertent movement or shaking of the user's hands while capturing such videos leads to corresponding shaking in the captured video.

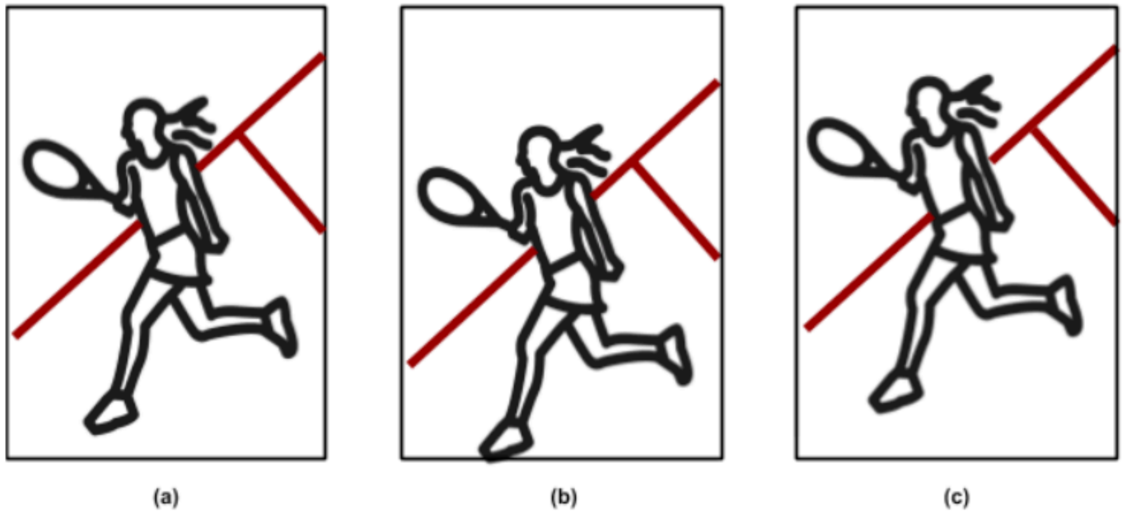

Fig. 1: Movement in captured video due to shaking

Fig. 1 illustrates an example of shaking in captured video due to shaking. Three frames of the video, captured in close time proximity, are shown in Fig. 1(a)-(c). As can be seen, while the scene is relatively unchanged, the tennis player appears at slightly different vertical positions in the scene. In Fig. 1(b) the foot of the tennis player is closer to the bottom of the video frame than in Fig. 1(a), while in Fig. 1(c) the foot is further away from the bottom than in Fig. 1(a). This sort of change can occur when the shaking of the hand is in a vertical direction that results in upward or downward motion of the camera during scene capture. Other types of motion, e.g., horizontal motion, or motion towards or away from the scene (change in depth) can also introduce shaking in the scene.

A video can be transformed into a 3D video to be rendered in a virtual reality (VR) environment, e.g., using machine learning techniques. After the transformation into 3D, the camera position is the origin of the 3D world. Correspondingly, the shaking in the scene that is due to camera motion at the time of capture manifests as a shake of the virtual world. This can cause the viewer of the VR environment to experience a roller coaster feeling since the scene shakes even when the VR user's head isn't actually moving.

DESCRIPTION

This disclosure describes techniques to deshake a 3D video for rendering in a virtual reality environment based on a captured 2D video to remove shaking that causes the roller coaster feeling.

Different points in a captured 2D video are tracked across a plurality of frames of the video. Tracking points are identified in a first frame of the video. For example, such points can include corners in the image or can be identified based on localizing the texture. The points are then tracked over the plurality of frames of the video to obtain trajectories of points. The trajectory can indicate the relative position of the point in sequential frames of the video.

At the time of constructing the 3D video, a depth map that describes each pixel by its depth is recovered for each video frame. Upon such recovery, each pixel has a 3D world space coordinate that is used to display the video in virtual reality. Each pixel is transformed into world space for the virtual environment by moving it out by its depth and landing it at a corresponding 3D location in the camera coordinate system. The trajectories of points in the 2D video is then used to set up an optimization to transform the 3D coordinate system such that the tracked points are stable (appear at the same position across multiple frames, when there is no change in the scene). For example, rotation and translation operations can be used to ensure that 3D points that correspond to the tracked 2D points are stable in the 3D space.

To perform such optimization, any type of continuous global optimization technique can be used. Based on the known correspondence between points in consecutive frames and associated depth, a rotation matrix and a translation vector is determined such that the 3D coordinates for corresponding points in consecutive frames are identical. The optimization can be performed jointly for both frames.

The described techniques can be used to render video in any type of virtual reality environment. The techniques are useful in VR environments that offer six degrees of freedom, where the rotation and movement of the user's head is tracked and the displayed scene is adjusted accordingly.

CONCLUSION

This disclosure describes techniques to track points across multiple frames of a 2D video. When the video is transformed into 3D, a 3D coordinate is established for each pixel of the video. The tracked points are used as input to optimization techniques that determine rotation and/or translation operations to transform the coordinate system such that the 3D world is stabilized.